OM nucleic - nucleic search, using sw model

```
          GenCore version 4.5
     Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: August 19, 2002, 12:44:21 ; Search time 6219.51 Seconds
                                           (without alignments)
                                           2644.623 Million cell updates/sec

Title: US-10-003-759-3
Perfect score: 786
Sequence: 1 atgaacgtcatgcgtgcggt........tttccgtaacggtgcagtag 786

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
          1:   gb_ba:*
          2:   gb_htg:*
          3:   gb_in:*
          4:   gb_om:*
          5:   gb_ov:*
          6:   gb_pat:*
          7:   gb_ph:*
          8:   gb_pl:*
          9:   gb_pr:*
         10:   gb_ro:*
         11:   gb_sts:*
         12:   gb_sy:*
         13:   gb_un:*
         14:   gb_vi:*
         15:   em_ba:*
         16:   em_fun:*
         17:   em_hum:*
         18:   em_in:*
         19:   em_mu:*
         20:   em_om:*
         21:   em_or:*
         22:   em_ov:*
         23:   em_pat:*
         24:   em_ph:*
         25:   em_pl:*
         26:   em_ro:*
         27:   em_sts:*
         28:   em_un:*
         29:   em_vi:*
         30:   em_htg_hum:*
         31:   em_htg_inv:*
         32:   em_htg_other:*
         33:   em_htgo_inv:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| Result No. | Score | Query Match | Length | DB ID | Description |
|---|---|---|---|---|---|
| 1 | 786 | 100.0 | 786 | AX339682 | 6 | AX339682 Sequence |
| 2 | 786 | 100.0 | 1877 | AX339680 | 6 | AX339680 Sequence |
| 3 | 746.8 | 95.0 | 1874 | RMU72637 | 1 | U72637 Rhodothermu |
| 4 | 131.8 | 16.8 | 1116 | AF233376 | 1 | AF233376 Streptomy |
| 5 | 131.8 | 16.8 | 1116 | AR129926 | 6 | AR129926 Sequence |
| 6 | 131.8 | 16.8 | 1116 | AR168360 | 6 | AR168360 Sequence |
| 7 | 105.8 | 13.5 | 1716 | AR168362 | 6 | AR168362 Sequence |
| 8 | 97.8 | 12.4 | 3527 | SHU51222 | 1 | U51222 Streptomyce |
| 9 | 96.2 | 12.2 | 1470 | SREGLS | 1 | X73953 S.rochei eg |
| 10 | 89 | 11.3 | 1311 | AF130408 | 1 | AF130408 Streptomy |
| 11 | 87.6 | 11.1 | 1565 | SLU04629 | 1 | U04629 Streptomyce |
| 12 | 87.6 | 11.1 | 41782 | SCG11A | 1 | AL133210 Streptomy |
| 13 | 69 | 8.8 | 11202 | AE005123 | 1 | AE005123 Halobacte |
| 14 | 68.8 | 8.8 | 4262 | AF335723 | 1 | AF335723 Burkholde |
| 15 | 64.4 | 8.2 | 14483 | AE006992 | 1 | AE006992 Mycobacte |
| 16 | 64.4 | 8.2 | 67200 | MTV017 | 1 | AL021897 Mycobacte |
| 17 | 57.6 | 7.3 | 1983 | AX047403 | 6 | AX047403 Sequence |
| 18 | 55 | 7.0 | 36368 | SC9B5 | 1 | AL035206 Streptomy |
| 19 | 54.8 | 7.0 | 346897 | AP002995 | 1 | AP002995 Mesorhizo |
| 20 | 54.6 | 6.9 | 45313 | SCD95A | 1 | AL357432 Streptomy |
| 21 | 54.6 | 6.9 | 125020 | AF429315 | 9 | AF429315 Homo sapi |
| 22 | 54 | 6.9 | 38995 | SCF34 | 1 | AL109974 Streptomy |
| 23 | 54 | 6.9 | 34182 | SCI11 | 1 | AL096849 Streptomy |
| 24 | 53.6 | 6.8 | 77534 | AF235504 | 1 | AF235504 Streptomy |
| 25 | 53 | 6.7 | 33779 | SCGD3 | 1 | AL096022 Streptomy |
| 26 | 52.8 | 6.7 | 125020 | AF429315 | 9 | AF429315 Homo sapi |
| 27 | 52.2 | 6.6 | 1849 | SGSTRB1F | 1 | X78973 S.galbus DS |
| 28 | 52 | 6.6 | 349116 | AP003003 | 1 | AP003003 Mesorhizo |
| 29 | 51.4 | 6.5 | 3644 | AF034576 | 5 | AF034576 Gallus ga |
| 30 | 51.2 | 6.5 | 22115 | SCC22 | 1 | AL096839 Streptomy |
| 31 | 50.6 | 6.4 | 5525 | SGSTRB1FG | 1 | X78974 S.glaucesce |
| 32 | 50.6 | 6.4 | 14596 | AE007168 | 1 | AE007168 Mycobacte |
| 33 | 50.6 | 6.4 | 25459 | SGAJ6985 | 1 | AJ006985 Streptomy |
| 34 | 50.6 | 6.4 | 30302 | MTCY6G11 | 1 | Z92774 Mycobacteri |
| 35 | 50.6 | 6.4 | 40806 | MSGY23 | 1 | AD000016 Mycobacte |
| 36 | 50 | 6.4 | 29625 | SCH22A | 1 | AL159178 Streptomy |
| 37 | 50 | 6.4 | 36394 | SCD35 | 1 | AL160312 Streptomy |
| 38 | 50 | 6.4 | 40104 | SCF11 | 1 | AL132662 Streptomy |
| 39 | 49.8 | 6.3 | 36734 | SC6G10 | 1 | AL049497 Streptomy |
| 40 | 49.6 | 6.3 | 26555 | SC2G1 | 1 | AL391014 Streptomy |
| 41 | 49.6 | 6.3 | 146555 | OSJN00013 | 2 | OSJN00013 Oryza sat |
| 42 | 49.6 | 6.3 | 215050 | AL646057 | 1 | AL646057 Ralstonia |
| 43 | 49.4 | 6.3 | 150814 | AP003541 | 1 | AP003541 Oryza sat |
| 44 | 49.2 | 6.3 | 12130 | AE005133 | 2 | AE005133 Halobacte |
| 45 | 49.2 | 6.3 | 41173 | SC141 | 1 | AL132648 Streptomy |

ALIGNMENTS

```
RESULT 1
AX339682
LOCUS          AX339682                  786 bp      DNA        linear   PAT 10-JAN-2002
DEFINITION     Sequence 3 from Patent WO0196382.
ACCESSION      AX339682
VERSION        AX339682.1   GI:18135584
KEYWORDS       .
SOURCE         Rhodothermus marinus.
  ORGANISM     Rhodothermus marinus
               Bacteria: CFB group; Rhodothermus group; Rhodothermus.
REFERENCE      1 (sites)
  AUTHORS      Wicher,K.B., Holst,O.P., Hachem,M.Y., Karlsson,E.M. and
               Hreggvidsson,G.O.
  TITLE        Thermostable cellulase
  JOURNAL      Patent: WO 0196382-A 3 20-DEC-2001;
               Prokaria ehf. (IS)
  FEATURES             Location/Qualifiers
               source          1..786
                               /organism="Rhodothermus marinus"
                               /db_xref="taxon:29549"
BASE COUNT      141 a    236 c    279 g    130 t
ORIGIN
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      August 15, 2002, 10:46:46 ; Search time 53.28 Seconds
                                         (without alignments)
                                         544.112 Million cell updates/sec

Title:       US-10-003-759-2
Perfect score: 1446
Sequence:    1 MNVNRAVLVLSLLLFGCDW...........FELWEGGAGLRSADFSVTVQ 261

Scoring table: BLOSUM62
             Gapop 10.0 , Gapext 0.5

Searched:    747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters:    747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_032802:*
         1:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
         2:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
         3:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
         4:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
         5:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
         6:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
         7:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
         8:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
         9:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
        10:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
        11:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
        12:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
        13:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
        14:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
        15:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
        16:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
        17:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
        18:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
        19:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
        20:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
        21:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
        22:  /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	% Match	Length	DB	ID	Description
1	1318.5	91.2		260	20	AAY06369	Rhodothermus marin
2	1318.5	91.2		260	21	AAB14882	Emericella deserto
3	1318.5	91.2		260	21	AAY84337	Amino acid sequence
4	470.5	32.5		105	20	AAY06342	Rhodothermus marin
5	400.5	27.7		386	21	AAY67497	Protien sequence o
6	396.5	27.4		371	20	AAY06367	Streptomyces sp. E
7	396.5	27.4		371	20	AAY08473	Actinomycete sp. 3
8	396.5	27.4		371	21	AAB14880	Streptomyces livid
9	396.5	27.4		371	21	AAY84345	Amino acid sequenc
10	396.5	27.4		371	21	AAY67496	Actinomycetes cell
11	343	23.7		381	20	AAY06368	Streptomyces livid
12	343	23.7		381	21	AAB14881	Rhodothermus marin
13	330.5	22.9		429	21	AAY84346	Amino acid sequence
14	255.5	17.7		261	20	AAW08462	Bacillus lichenifo
15	252.5	17.5		264	20	AAY06370	Erwinia carotovora
16	252.5	17.5		264	21	AAB14883	Erwinia carotovora
17	252.5	17.5		264	21	AAY84348	Amino acid sequenc
18	224.5	15.5		259	20	AAY06351	Aspergillus aculea
19	224.5	15.5		259	21	AAB14864	Aspergillus aculea
20	224.5	15.5		259	20	AAY84329	Amino acid sequence
21	214	14.8		239	20	AAY06353	Aspergillus kawach
22	214	14.8		239	21	AAB14866	Aspergillus kawach
23	214	14.8		239	21	AAY84331	Amino acid sequenc
24	213.5	14.8		254	21	AAB14869	Humicola insolens
25	213.5	14.8		254	21	AAY84334	Amino acid sequenc
26	210.5	14.6		246	20	AAY06366	Emericella deserto
27	210.5	14.6		246	21	AAB14879	Actinomycete 11AG8
28	210.5	14.6		246	21	AAY84344	Amino acid sequence
29	210.5	14.6		254	21	AAB14868	Humicola grisea EG
30	210.5	14.6		254	21	AAY84333	Amino acid sequence
31	210.5	14.6		254	22	AAU07558	Humicola grisea en
32	210.5	14.6		254	22	AAU07584	Humicola grisea en
33	210	14.6		254	22	AAU07583	Humicola grisea en
34	203	14.0		255	20	AAY06356	Humicola insolens
35	200.5	13.9		348	20	AAY06362	Gliocladium roseum
36	200.5	13.9		348	21	AAB14875	Gliocladium roseum
37	200.5	13.9		348	21	AAY84340	Amino acid sequence
38	199	13.4		104	20	AAY06337	Streptomyces sp. E
39	194	13.4		253	21	AAY06355	Humicola grisea EG
40	191.5	13.2		234	22	AAU07582	Trichoderma reesel
41	186.5	12.9		244	19	AAW68593	Tiarosporella phas
42	186.5	12.9		244	20	AAY44341	T. phaseolina xylo
43	186	12.9		244	20	AAY06359	Fusarium javanicum
44	186	12.9		244	21	AAB14872	Fusarium javanicum
45	186	12.9		244	21	AAY84337	Amino acid sequenc

ALIGNMENTS

```
RESULT 1
AAY06369

ID   AAY06369 standard; Protein; 260 AA.
XX
AC   AAY06369;
XX
DT   06-SEP-1999  (first entry)
XX
DE   Rhodothermus marinus EGIII-like cellulase.
XX
KW   Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
KW   food processing; grain wet milling; pulp; paper.
XX
OS   Rhodothermus marinus.
XX
XX
PN   WO9910965-A2;
XX
PD   1999-06-16;
XX
PF   14-DEC-1998;  98WO-US26552.
XX
PR   16-DEC-1997;  97US-0991720.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Bower BS, Fowler T, Phillips JI;
XX
DR   WPI; 1999-395187/33.
XX
PT   EGIII like cellulase
XX
PS   Example; Fig 6; 47pp; English.
XX
CC   The present polypeptide represents a full-length sequence of a
```

```
CC   novel EGIII-like cellulase of Rhodothermus marinus.  It was
CC   deduced from a gene sequence isolated from genomic DNA using PCR
CC   primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC   of Trichoderma reesel EGIII cellulase and related enzymes.  PCR
CC   has been used to identify novel EGIII-like enzymes, including the
CC   present protein, from bacterial and fungal sources (see AAY06331-70).
CC   The sequence shows homology to T. reesel EGIII (see AAY06330).  Also
CC   provided by the invention are vectors, host cells and methods
CC   for the recombinant production of such enzymes, which can be used
CC   in the treatment of cellulose-containing textiles, as feed
CC   additives, in the treatment of wood pulp, in the reduction of
CC   biomass to glucose, in the stone washing of indigo dyed denim, or
CC   as laundry detergent components (all claimed).
XX
SQ   Sequence 260 AA;

Query Match            91.2%;  Score 1318.5;  DB 20;  Length 260;
Best Local Similarity  90.3%;  Pred. No. 1.7e-118;
Matches   241;  Conservative    4;  Mismatches    9;  Indels   13;  Gaps    2;

Qy           1 MNVMRAVLVLSLLLFGCDWLFPDGDNGKEPEPEPTVELCGRWDARDVAGGRYRVINN    60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 mnvmravlvlsllllfgcdwlfpdgdngkepepeptvelcgrwdardvaggryrvinn   60

Qy          61 VWGAETAQCIEVGLETGNFTITRADHDNGNNVAAYPAIYFGCHWGA------CTSNSGLP  114
               ||||||||||||||||||||||||||||||||||||||||        :  : :|
Db          61 vwgaetaqclevgletgnftitradhdngnnvaaypalyfgchwaparairdcaaragav  120

Qy         115 RRVQELSDVRTSWTLTPITTGRWNAAYDIWFSPVTNSGNGYSGGAELMIWLNWNGGVMPG  174
               ||   ||   :::::||||||||||||||||||||||||||||||||
Db         121 rraheld------vtpltgrwnaaydiwfspvtnsgngysggaelmiwlnwnggvmpg   173

Qy         175 GSRVATVELAGATWEVWYADWDWNYIAYRRTPTTSVSELDLKAFIDDAVARGYIRPEWY  234
               |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db         174 gsrvatvelagatwevwyadwdwnyiayrrtptisvseldikafiddavargyirpewy  233

Qy         235 LHAVETGFELWEGGAGLRSADFSVTVQ 261
               ||||||||||||||||:||||||||
Db         234 lhavetgfelweggaglrtadfsvtvq 260

RESULT 2
AAB14882
ID   AAB14882 standard; Protein; 260 AA.
XX
AC   AAB14882;
XX
DT   21-NOV-2000  (first entry)
XX
DE   Emericella desertoru EGIII-like cellulase.
XX
KW   Emericella desertoru; Trichoderma reesel; endoglucanase III; EGIII;
KW   cellulase; mutant; enzyme stability; textile treatment;
KW   wood pulp treatment; feed additive; detergent.
XX
OS   Emericella desertoru.
XX
PN   WO200037614-A2.
XX
PD   29-JUN-2000.
XX
PF   12-NOV-1999;  99WO-US26704.
XX
PR   18-DEC-1998;  98US-0216295.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Mitchinson C, Wendt DJ;
XX
DR   WPI; 2000-482483/42.
XX
PT   Novel endoglucanase III or endoglucanase III-like cellulase useful for
```

```
PT   treating textiles and wood pulp comprises a substitution or deletion at
PT   specified positions in the wild form of endoglucanase III -
XX
PS   Example 1; Fig 3; 52pp; English.
XX
CC   The present sequence is a cellulase related to endoglucanase III (EGIII)
CC   from Trichoderma reesel. EGIII-like genes were isolated from genomic DNA
CC   libraries constructed from various microorganisms by PCR. The isolated
CC   genes showed significant homology to EGIII from T. reesel. Certain
CC   subsitution and deletion mutations have been incorporated into EGIII and
CC   EGIII-like cellulases to produce variant enzymes with improved stability,
CC   e.g. increased resistance to temperature stress. The mutants may be used
CC   in textile and wood pulp treatment, as a feed additive, and for reducing
CC   biomass to glucose. They are also useful for stonewashing or indigo dyed
CC   denim and as an agent in laundry and dish detergents.
XX
SQ   Sequence 260 AA;

Query Match            91.2%;  Score 1318.5;  DB 21;  Length 260;
Best Local Similarity  90.3%;  Pred. No. 1.7e-118;
Matches   241;  Conservative    4;  Mismatches    9;  Indels   13;  Gaps    2;

Qy           1 MNVMRAVLVLSLLLFGCDWLFPDGDNGKEPEPEPTVELCGRWDARDVAGGRYRVINN    60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 mnvmravlvlsllllfgcdwlfpdgdngkepepeptvelcgrwdardvaggryrvinn   60

Qy          61 VWGAETAQCIEVGLETGNFTITRADHDNGNNVAAYPAIYFGCHWGA------CTSNSGLP  114
               ||||||||||||||||||||||||||||||||||||||||        :  : :|
Db          61 vwgaetaqclevgletgnftitradhdngnnvaaypalyfgchwaparairdcaaragav  120

Qy         115 RRVQELSDVRTSWTLTPITTGRWNAAYDIWFSPVTNSGNGYSGGAELMIWLNWNGGVMPG  174
               ||   ||   :::::||||||||||||||||||||||||||||||||
Db         121 rraheld------vtpltgrwnaaydiwfspvtnsgngysggaelmiwlnwnggvmpg   173

Qy         175 GSRVATVELAGATWEVWYADWDWNYIAYRRTPTTSVSELDLKAFIDDAVARGYIRPEWY  234
               |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db         174 gsrvatvelagatwevwyadwdwnyiayrrtptisvseldikafiddavargyirpewy  233

Qy         235 LHAVETGFELWEGGAGLRSADFSVTVQ 261
               ||||||||||||||||:||||||||
Db         234 lhavetgfelweggaglrtadfsvtvq 260

RESULT 3
AAY84347
ID   AAY84347 standard; Protein; 260 AA.
XX
AC   AAY84347;
XX
DT   12-JUL-2000  (first entry)
XX
DE   Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.
XX
KW   Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
KW   cellulase; textile processing; textile cleaning; stonewashing;
KW   indigo dyed denim; cellulose containing fabric; fabric smoothness;
KW   pill removal; fibril removal; cotton; cellulosic fibre; dying; detergent;
KW   animal feed; wood pulp; paper; grain; biomass reduction; glucose.
XX
OS   Rhodothermus marinus.
XX
PN   WO200014208-A1
XX
PD   16-MAR-2000.
XX
PF   24-AUG-1999;  99WO-US19154.
XX
PR   03-SEP-1998;  98US-0146729.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Fowler T;
```

```
XX
DR   WPI; 2000-271052/23.
XX
PT   Novel variant endoglucanase III-like cellulases with improved
PT   surfactant stability and resistance to temperature stress, useful for
PT   textile processing or cleaning, treating wood pulp, food and grain, and
PT   reducing biomass to glucose .
XX
PS   Disclosure; Page 66-67; 73pp; English.
XX
CC   The present sequence represents an endoglucanase III (EGIII)-like
CC   cellulase. The cellulase has homology to the Trichoderma reesei EGIII
CC   protein. The variant cellulases have improved temperature stability,
CC   and improved surfactant stability. The variant cellulases and
CC   compositions containing them are used in textile processing or cleaning,
CC   e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC   or appearance of cellulose containing fabrics (e.g. improving fabric
CC   smoothness or removing pills and fibrils). The compositions may also be
CC   used for the removal of immature or dead cotton from cellulosic fibres
CC   or fabric, which can cause uneven dying. The cellulase may also be used
CC   in a detergent composition for washing laundry and dishes and in the
CC   treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC   The enzymes may also be used in the reduction of biomass to glucose.
XX
SQ   Sequence   260 AA;

Query Match       91.2%;    Score 1318.5; DB 21; Length 260;
Best Local Similarity  90.3%;    Pred. No. 1.7e-118;
Matches 241; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

Qy    1 MNVMRAVLVLSLLLFGCDWLFPDGDGKEPEPEPTVELCGRWDARDVAGGRYRVINN 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 mnvmravlvlslllfgcdwlfpdgdngkepepepeptvelcgrwdardvaggryrvinn 60

Qy   61 VRGAETAQCIEVGLETGNFTITRADHDNGNNVAAYPAIYFGCHWGA------CTSNSGLP 114
          |||||||||||||||||||||||||||||||||||||     |  : :|
Db   61 vvgaetaqclevgletgnftitradhdngnnvaaypaiyfgchwaparairdcaaragav 120

Qy  115 RRVQELSDVRTSWTLTPITTGRWNAAYDIWFSPVTNSGNGYSGGAELMIWLNWNGGVMPG 174
          ||     :::||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 rrraheld------vtpittgrwnaaydiwfspvtnsgngysggaelmiwlnwnggvmpg 173

Qy  175 GSRVATVELAGATWEVWYADWDWNYIAYRRTFPTTSVSELDLKAFIDDAVARGYIRPEWY 234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 gsrvatvelagatwevwyadwdwnylayrrttpttsvseldlkafiddavargylrpewy 233

Qy  235 LHAVETGFELWEGGAGLRSADFSVTVQ 261
          |||||||||||||||||||||||||||
Db  234 lhavetgfelweggaglrtadfsvtvq 260

RESULT 4
AAY06342
ID   AAY06342 standard; Protein; 105 AA.
XX
AC   AAY06342;
XX
DT   06-SEP-1999  (first entry)
XX
DE   Rhodothermus marinus EGIII-like cellulase (partial sequence).
XX
KW   Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
KW   food processing; grain wet milling; pulp; paper.
XX
OS   Rhodothermus marinus.
XX
PN   WO9931255-A2.
XX
PD   24-JUN-1999.
XX
PF   14-DEC-1998;   98WO-US26552.
XX
PR   16-DEC-1997;   97US-099720.
XX
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Bower BS, Fowler T, Phillips JI;
XX
DR   WPI; 1999-395187/33.
XX
PT   EGIII like cellulase
XX
PS   Example; Fig 3; 47pp; English.
XX
CC   The present polypeptide represents a partial sequence of a novel
CC   EGIII-like cellulase of Rhodothermus marinus. It was deduced from
CC   a partial gene sequence isolated from genomic DNA using PCR
CC   primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC   of Trichoderma reesei EGIII cellulase and related enzymes.  PCR
CC   has been used to identify novel EGIII-like enzymes, including the
CC   present polypeptide, from bacterial and fungal sources (see
CC   AAY06331-70). Also provided by the invention are vectors, host
CC   cells and methods for the recombinant production of such enzymes,
CC   which can be used in the treatment of cellulose-containing textiles,
CC   as feed additives, in the treatment of wood pulp, in the reduction
CC   of biomass to glucose, in the stone washing of indigo dyed denim,
CC   or as laundry detergent components (all claimed).
XX
SQ   Sequence   105 AA;

Query Match       32.5%;    Score 470.5; DB 20; Length 105;
Best Local Similarity  77.7%;    Pred. No. 1.5e-37;
Matches 87; Conservative 3; Mismatches 9; Indels 13; Gaps 2;

Qy   59 NNVWGAETAQCIEVGLETGNFTITRADHDNGNNVAAYPAIYFGCHWGA------CTSNSG 112
          ||||||||||||||||||||||||||||||||||||||     |  : :|
Db    1 nnvwgaetaqclevgletgnftitradhdngnnvaaypaiyfgchwaparairdcaaraag 60

Qy  113 LPRRVQELSDVRTSWTLTPITTGRWNAAYDIWFSPVTNSGNGYSGGAELMIW 164
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 avrraheld------vtpittgrwnaaydiwfspvtnsgngysggaelmiw 105

RESULT 5
AAY67497
ID   AAY67497 standard; Protein; 386 AA.
XX
AC   AAY67497;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Protein sequence of CelA and cellulase 11AG8 fusion.
XX
KW   Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
KW   pulp; paper; cellulase 11AG8.
XX
OS   Streptomyces lividans.
OS   Synthetic.
XX
FH   Key          Location/Qualifiers
FT   Peptide      1..46
FT                /note= "celA signal sequence"
FT   Protein      47..386
FT                /note= "cellulase 11AG8 mature peptide"
XX
PN   WO200009707-A1.
XX
PD   24-FEB-2000.
XX
PF   28-MAY-1999;   99WO-US11971.
XX
PR   24-JUN-1998;   98US-0104308.
PR   18-NOV-1998;   98WO-US24649.
PR   28-MAY-1999;   99US-0321981.
XX
```

ACCESSION NUMBER: 2001:454833   CAPLUS
DOCUMENT NUMBER: 135:45237
TITLE: **Deletion** of a cytotoxic, N-terminal putative
                           signal peptide results in a significant increase in
                           production yields in Escherichia coli and improved
                           specific activity of Cel12A from Rhodothermus marinus
AUTHOR(S):   Wicher, K. B.; Abou-Hachem, M.; Halldorsdottir, S.;
                           Thorbjarnadottir, S. H.; Eggertsson, G.; Hreggvidsson,
                           G. O.; Karlsson, E. Nordberg; Holst, O.
CORPORATE SOURCE: Department of Biotechnology, Center for Chemistry and
                           Chemical Engineering, Lund University, Lund, 221 00,
                           Swed.
SOURCE: Applied Microbiology and Biotechnology (2001), 55(5),
                           578-584
                           CODEN: AMBIDG; ISSN: 0175-7598
PUBLISHER: Springer-Verlag
DOCUMENT TYPE: Journal
LANGUAGE: English

AB   The thermostable **cellulase** Cel12A from Rhodothermus marinus was
     produced at extremely low levels when expressed in Escherichia coli and
     was cytotoxic to the cells. In addn., severe aggregation occurred when
     moderately high concns. of the enzyme were heat-treated at 65.degree.C,
     the growth optimum of R. marinus. Sequence anal. revealed that the
     catalytic module of this enzyme is preceded by a typical linker sequence
     and a highly hydrophobic putative signal peptide. Two **deletion**
     mutants lacking this hydrophobic region were cloned and successfully
     expressed in E. coli. These results indicated that the N-terminal
     putative signal peptide was responsible for the toxicity of the
     full-length enzyme in the host organism. This was further corroborated by
     cloning and expressing the **hydrophobic N-**
     terminal domain in E. coli, which resulted in extensive cell
     lysis. The **deletion** mutants, made up of either the catalytic
     module of Cel12A or the catalytic module and the putative linker sequence,
     were characterized and their properties compared to those of the
     full-length enzyme. The specific activity of the mutants was approx.
     three-fold higher than that of the full-length enzyme. Both mutant
     proteins were highly thermostable, with half-lives exceeding 2 h at
     90.degree.C and unfolding temps. up to 103.degree.C.